Diseño de experimentos - Experimentos de un factor

http://www.academia.utp.ac.pa/humberto-alvarez/diseno-de-experimentos-y-regresion

Introducción

- Aunque se sigue considerando un solo factor, se presentan los diseños experimentales que se utilizan cuando el objetivo es comparar más de dos tratamientos.
- El interés se centra no solo en comparar los tratamientos en cuanto a sus medias poblacionales, sino compararlos con respecto a sus varianzas.
- Se quiere decidir si los tratamientos son iguales estadísticamente en cuanto a sus medias, frente a la alternativa de que al menos dos de ellos sean diferentes.
- La hipótesis fundamental a probar cuando se comparan varios tratamientos es:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$$

$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j$$

Diseño completamente al azar y ANOVA

- Un diseño se llama *completamente al azar* porque todas las corridas experimentales se realizan en orden aleatorio completo.
- Así, durante el experimento las N pruebas se corren al azar, de manera que los posibles efectos ambientales y temporales se vayan repartiendo equitativamente entre los tratamientos.

Tratamientos				
T_1	T_2	T_3	…	T_k
Y_{11}	Y_{21}	Y_{31}	…	Y_{k1}
Y_{12}	Y_{22}	Y_{32}	…	Y_{k2}
Y_{13}	Y_{23}	Y_{33}	…	Y_{k3}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Y_{1n_1}	Y_{2n_2}	Y_{3n_3}	…	Y_{kn_k}

Aspectos teóricos

- Supóngase que se tienen k poblaciones o tratamientos, independientes y con medias desconocidas $\mu_1, \mu_2, \ldots, \mu_k$, así como varianzas desconocidas que se suponen iguales $\sigma^2{}_1=\sigma^2{}_2=\ldots=\sigma^2{}_k=\sigma^2$
- Las poblaciones pueden ser k métodos de producción, k tratamientos, k grupos, etc., y sus medias se refieren o son medidas en términos de la variable de respuesta.
- Se decide realizar un experimento completamente al azar para comparar las poblaciones, en principio mediante la hipótesis de igualdad de medias.

- El número de tratamientos *k* es determinado por el investigador y depende del problema particular de que se trata.
- El número de observaciones por tratamiento (*n*) debe escogerse con base en la variabilidad que se espera observar en los datos, así como en la diferencia mínima que el experimentador considera que es importante detectar.
- Con este tipo de consideraciones, por lo general se recomiendan entre 5 y 30 mediciones en cada tratamiento.
- Por ejemplo, se usa $n = 10$ cuando las mediciones dentro de cada tratamiento tienen un comportamiento consistente (con poca dispersión). En el otro extremo, se recomienda $n = 30$ cuando las mediciones muestran bastante dispersión.
- Cuando es costoso o tardado realizar las pruebas para cada tratamiento se puede seleccionar un número menor de repeticiones, con lo cual sólo se podrán detectar diferencias grandes entre los tratamientos.
- Cuando $n_i = n$ para toda i se dice que el *diseño es balanceado*.

- Las observaciones obtenidas de cada tratamiento se podrán describir con el modelo lineal estadístico:

$$Y_{ij}=\mu + \tau_i + \varepsilon_{ij}$$

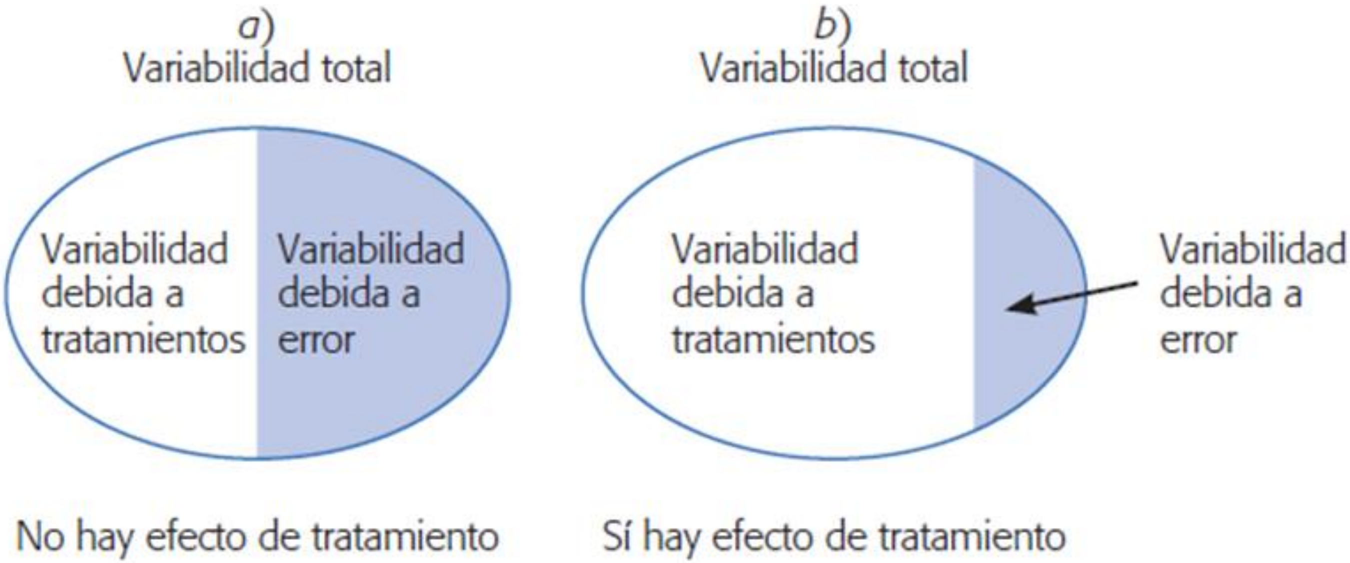
- Donde μ es la media global que es común a cada tratamiento, τ_i que es relativo al tratamiento i y ε_{ij} es el error atribuible a la medición Y_{ij}.
- Este modelo implica que en el diseño completamente al azar actuarían a lo más dos fuentes de variabilidad: los *tratamientos* y el *error aleatorio*.
- La media global μ de la variable de respuesta no se considera una fuente de variabilidad por ser una constante común a todos los tratamientos, que hace las veces de punto de referencia con respecto al cual se comparan las respuestas medias de los tratamientos.

Diferencias entre prueba “t” y ANOVA

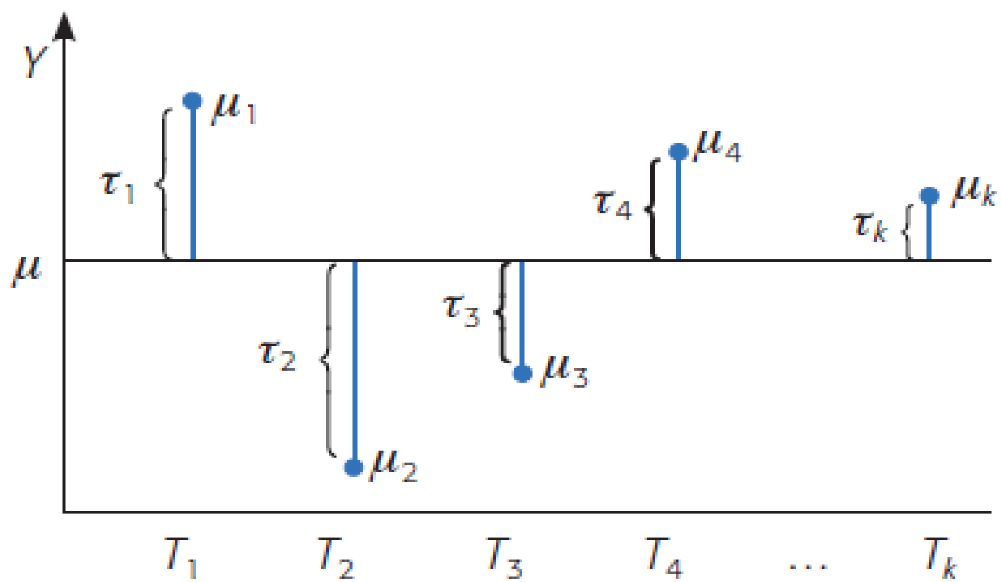
- El Test “t” es muy útil para muestras pequeñas y para establecer la diferencia entre las Medias Aritméticas de dos variables solamente.
- ANOVA amplía el rango del “t” test para determinar si las medias de varios grupos, no sólo de dos, son o no iguales.

Análisis de varianza (ANOVA) para el diseño completamente al azar

- El *análisis de varianza* (ANOVA) es la técnica central en el análisis de datos experimentales.
- Su objetivo es separar la variación total en las partes con las que contribuye cada fuente de variación en el experimento, en otras palabras la variabilidad debida a los tratamientos y al error.
- Cuando los tratamientos no dominan contribuyen igual o menos que el error, se concluye que las medias son iguales
- Cuando los tratamientos predominan “claramente” sobre el error, es cuando se concluye que los tratamientos tienen efecto y las medias son diferentes.



Variación total de los componentes en un diseño completamente aleatorio



Efecto de los tratamientos en un diseño completamente aleatorio

Hipótesis en ANOVA

- El objetivo del análisis de varianza en el DCA es probar la hipótesis de igualdad de los tratamientos con respecto a la media de la correspondiente variable de respuesta:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$$

$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j$$

- la cual se puede escribir en forma equivalente como:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0$$

$$H_A : \tau_i \neq 0 \text{ para algún } i$$

- Si se acepta H_0} se confirma que los efectos sobre la respuesta de los *k* tratamientos son estadísticamente nulos y en caso de rechazar se estaría concluyendo que al menos un efecto es diferente de cero.

Notación de puntos:

$Y_{i\bullet}$ = Suma de las observaciones del tratamiento i.
$\overline{Y}_{i\bullet}$ = Media de las observaciones del i-ésimo tratamiento.
$Y_{\bullet\bullet}$ = Suma total de las $N = n_1 + n_2 + \ldots + n_k$ mediciones.
$\overline{Y}_{\bullet\bullet}$ = Media global o promedio de todas las observaciones.

$$Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij}; \overline{Y}_{i\bullet} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}; Y_{\bullet\bullet} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij} \quad N = \sum_{i=1}^{k} n_i$$

$$\overline{Y}_{\bullet\bullet} = \frac{Y_{\bullet\bullet}}{N}; \; i = 1, 2, \ldots, k$$

$$SC_T = \sum_{i=1}^{k} n_i (\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\bullet})^2$$

donde el primer componente es la *suma de cuadrados de tratamientos* (SC_{TRAT}) y el segundo es la *suma de cuadrados del error* (SC_E).

Tabla de ANOVA para un diseño completamente al azar

FV	SC	GL	CM	F_0	Valor-p
Tratamientos	$SC_{TRAT} = \sum_{i=1}^{k} \frac{Y_{i\bullet}^2}{n_i} - \frac{Y_{\bullet\bullet}^2}{N}$	$k-1$	$CM_{TRAT} = \frac{SC_{TRAT}}{k-1}$	$\frac{CM_{TRAT}}{CM_E}$	$P(F > F_0)$
Error	$SC_E = SC_T - SC_{TRAT}$	$N-k$	$CM_E = \frac{SC_E}{N-k}$		
Total	$SC_T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{\bullet\bullet}^2}{N}$	$N-1$			

ANOVA supone que la variable de respuesta se distribuye *normal*, con *varianza constante* (los tratamientos tienen varianza similar), y que las mediciones son *independientes* entre sí. Estos supuestos deben verificarse para estar más seguros de las conclusiones obtenidas.

La prueba de diferencia mínima significativa LSD

- ANOVA tiene como objetivo probar la hipótesis

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$$

$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j$$

- Se podría pensar que una forma de probar la hipótesis nula es mediante pruebas *T* de Student aplicadas a todos los posibles pares de medias.
- Esta requiere de $\frac{k(k-1)}{2}$ comparaciones, donde k es la cantidad de tratamientos.
- Una alternativa es utilizar la Diferencia Mínima Significativa (LSD).
- Es la diferencia mínima que debe haber entre dos medias muestrales para considerar que dos tratamientos son diferentes.

- En el LSD, el estadístico de prueba para cada hipótesis es la correspondiente diferencia en el valor absolutos entre sus medias muestrales, tal que se rechaza la hipótesis H_0 si:

$$\left|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\right| > t_{\alpha/2,\,N-k}\sqrt{CM_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} = LSD$$

- Si el diseño es balanceado, en otras palabras, si $n_1 = n_2 = \ldots = n_k = n$

$$LSD = t_{\alpha/2,\ N-k}\sqrt{2CM_E/n}$$

Otras pruebas posteriores de comparación de medias

- **Método de Tukey**: es un método más conservador que el LSD.
- Consiste en comparar medias muestrales con el valor crítico dado por:

$$T_\alpha = q_\alpha\,(k, N-k)\ \sqrt{CM_E/n_i}$$

- Donde CM_E es el cuadrado medio del error, n es el número de observaciones, N-k son los grados de libertad para el error, α es el nivel de significancia y $q_\alpha(k, N\text{-}k)$ es un estadístico estandarizado de puntos porcentuales.

Otras pruebas posteriores de comparación de medias

- **Método de Duncan**: En este caso, si las k muestras son de igual tamaño, los k promedios se acomodan en orden ascendente y el error estándar de los promedios se determina como:

$$s_{\bar{Y}_{i\cdot}} = \sqrt{\frac{CM_E}{n}}$$

- Con los k-1 grados de libertad se obtienen los rangos de significancia $R_p = r_\alpha(p, l)S_{\bar{y}}$, para todo p = 2, 3, ..,k

Ejemplo

- Se tienen que probar cuatro posibles tratamientos sobre cierto material. La dureza del material después de los procesos se da a continuación. Se desea saber si hay diferencias entre los tratamientos.:

A	B	C	D
6	7	11	10
8	9	16	12
7	10	11	11
8	8	13	9

Caso	Tratamiento	Dureza
1	A	6
2	A	8
3	A	7
4	A	8
5	B	7
6	B	9
7	B	10
8	B	8
9	C	11
10	C	16
11	C	11
12	C	13
13	D	10
14	D	12
15	D	11
16	D	9

Análisis de la varianza

Variable	N	R^2	R^2 Aj	CV
Dureza	16	0.70	0.63	16.08

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo.	69.50	3	23.17	9.42	0.0018
Tratamiento	69.50	3	23.17	9.42	0.0018
Error	29.50	12	2.46		
Total	99.00	15			

Test:LSD Fisher Alfa=0.05 DMS=2.41560

Error: 2.4583 gl: 12

Tratamiento	Medias	n	E.E.			
C	12.75	4	0.78	A		
D	10.50	4	0.78	A	B	
B	8.50	4	0.78		B	C
A	7.25	4	0.78			C

Medias con una letra común no son significativamente diferentes (p > 0.05)

Test:Tukey Alfa=0.05 DMS=3.29156

Error: 2.4583 gl: 12

Tratamiento	Medias	n	E.E.		
C	12.75	4	0.78	A	
D	10.50	4	0.78	A	B
B	8.50	4	0.78		B
A	7.25	4	0.78		B

Medias con una letra común no son significativamente diferentes (p > 0.05)

Test:Duncan Alfa=0.05

Error: 2.4583 gl: 12

Tratamiento	Medias	n	E.E.			
C	12.75	4	0.78	A		
D	10.50	4	0.78	A	B	
B	8.50	4	0.78		B	C
A	7.25	4	0.78			C

Medias con una letra común no son significativamente diferentes (p > 0.05)

Prueba T para muestras Independientes

Clasific	Variable	Grupo 1	Grupo 2	n(1)	n(2)	Media(1)	Media(2)	Media(1)-Media(2)	LI(95)	LS(95)	pHomVar	T	p-valor	prueba
Tratamiento	Dureza	{A}	{B}	4	4	7.25	8.50	-1.25	-3.22	0.72	0.6356	-1.56	0.1708	Bilateral
Tratamiento	Dureza	{A}	{C}	4	4	7.25	12.75	-5.50	-8.62	-2.38	0.1720	-4.31	0.0050	Bilateral
Tratamiento	Dureza	{A}	{D}	4	4	7.25	10.50	-3.25	-5.22	-1.28	0.6356	-4.04	0.0068	Bilateral
Tratamiento	Dureza	{B}	{C}	4	4	8.50	12.75	-4.25	-7.54	-0.96	0.3473	-3.16	0.0196	Bilateral
Tratamiento	Dureza	{B}	{D}	4	4	8.50	10.50	-2.00	-4.23	0.23	>0.9999	-2.19	0.0710	Bilateral
Tratamiento	Dureza	{C}	{D}	4	4	12.75	10.50	2.25	-1.04	5.54	0.3473	1.67	0.1457	Bilateral